Experimental Linguistics in the Field:
Opportunities, Challenges, Resources, and Ethical Issues

Sonja Eisenbeiss

University of Essex

seisen@essex.ac.uk

Revised version, October 2013-10-13

This is a very initial draft. Comments are very welcome, but please do not quote without permission.
Experimental Linguistics in the Field: Opportunities, Challenges, Resources, and Ethical Issues

Sonja Eisenbeiss
University of Essex
seisen@essex.ac.uk

Revised version, October 2013-10-13

Introduction
Opportunities and how to Make the Best Use of them
   Extending Linguistic Experimentation to Under-Researched Languages
   Extending Psycholinguistic Experimentation to Under-Researched Populations
   Enriching Linguistic Fieldwork and Descriptive Linguistics with Experimental Methods
   Making Use of Opportunities
Dealing with the Lack of a Fully Equipped Laboratory
   Equipment
   Software
   Practical Issues
Dealing with a Lack of Psycholinguistic Resources
   Corpora, Lexical Information and Frequency Measures
   Relatedness Measures for Word Pairs
   Practical Issues
Working with Participants from Different Cultural Backgrounds
   Stimulus Materials
   Tasks, Experimental Designs, and Practical Issues
Adapting "Traditional" Experimental Tasks to "Non-Traditional Contexts
   Tasks for the Study of Lexical Representations and Processing
   Tasks for the Study of Grammatical Representations and Processing
Ethical Issues
   Transparency, Quality, and Informed Consent
   Confidentiality, Anonymity and the Data Protection Act

Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc
Introduction

Experimental linguists study diverse linguistic phenomena and populations – children, and adults, populations with or without language disorders, monolingual and multi-lingual speakers and different types of language learners. They use a variety of experimental paradigms, which all share the strict control of procedures and stimulus items, but involve a broad range of production, comprehension, and judgment tasks (for overviews see e.g. Blom & Unsworth, 2010; Cowart, 1997; MacDaniel, MacKee, & Cairns, 1996, Menn & Ratner, 1999, Myers, 2009, Sekerina, Fernández, & Clahsen, 2008; Wei & Moyer, 2008; for updates, see http://experimentalfieldlinguistics.wordpress.com/readings/).

Despite this variety in populations and experimental paradigms, most experimental linguists have quite similar working conditions; and they tend to take them for granted until they go beyond their laboratory and their usual sets of languages and participants: They often have a more or less well-equipped laboratory with specialist equipment and commercial, specialised software and hardware for reaction-time (RT) measurements. Many experimental linguists also have direct or indirect access to laboratories for eye-movement or brain-imaging measurements. Moreover, even experimental linguists that carry out cross-linguistic studies tend to work on well-described languages for which they can quite easily find research resources like grammars, dictionaries, lexical databases and sometimes even pre-tested stimuli with naming accuracy and response-time information (for links and references, see http://experimentalfieldlinguistics.wordpress.com/experimental-materials/ and section ##).

Finally, experimental linguists typically work with a comparatively homogeneous pool of (computer-)literate participants who come from the same culture as the researcher. In studies on
language processing in adults, participants are usually university students that have either taken part in experiments themselves before or are at least familiar with the concept of experiments. In "traditional" studies with young children, child participants are usually illiterate and may not always have previous experience with experiments. However, researchers and child participants typically share a language and culture; and many researchers work with schools or day care centres where parents, staff and children are used to researchers coming in with their "games" on a more or less regular basis.

Experimental studies in such settings have produced a wealth of data and insights into our ability to acquire and process language (for overviews, see e.g. Altmann, 1997; Altmann & Gaskell, 2007; Eisenbeiss (subm.), Kess, 1992; Traxler, 2011; Traxler & Gernsbacher, 2011). However, more and more researchers have become wary of claims about universals of language acquisition and processes when most available psycholinguistic studies are based on a limited set of languages and researchers tend to limit themselves to work with often quite highly educated and (computer)literate participant groups (for discussion, see e.g. Jaeger & Norcliffe, 2009). Hence, a growing number of experimental linguists is moving beyond their "comfort zone" to (i) carry out experiments outside a traditional laboratory environment, (ii) study under-researched languages with limited or non-existent research resources, or (iii) work with populations that are less familiar with experiments and computers than the typical participants of linguistic experiments. These researchers often find it difficult to convince their funders, institutions and (fieldworking) colleagues of the unique opportunities that experimental field linguistics can offer. Hence, section # will present different types of studies that have demonstrated the special contribution that experimental linguistics in the field can make to studies into the human capacity for language acquisition and processing. Opportunities rarely come without a price, though; and
Experimental linguists quickly become aware of the particular challenges that are associated with leaving the "traditional" laboratory setting. These challenges will be discussed in sections ###, where I will also present resources and tools that can help researchers to overcome them. Section # focuses on the challenges that result from the lack of a fully equipped lab, while section ## addresses the particular difficulties involved in work with under-researched languages. The focus of section ## is on the constraints for stimulus and task selection that arise in cross-cultural studies and in studies with "non-traditional" participant populations. Finally, I will discuss the particular ethical issues in situations where researches must meet both the ethics standards for fieldworkers and the ethics standards for experimental studies with human participants - while they are also dealing with new ethical issues that are associated with experimental work in a field setting (section #).

**Opportunities and how to Make the Best Use of them**

When experimental linguists study the way in which our linguistic knowledge is organised, acquired, and used in language processing, they often develop models that are supposed to capture universals of linguistic knowledge, language acquisition and language processing mechanisms. However, the majority of studies that have evaluated these models are based on a very small sample of languages that share many lexical and typological properties, with a continuing strong focus on English and other closely related European languages. In the following sections, I will show that some common claims about linguistic and processing universals and the way in which languages vary have been challenged by research that goes beyond the range of the languages and populations that experimental linguists typically investigate (sections # and #). Moreover, I will argue that taking the experimental approach to...
the field may also improve the quality of language descriptions and hence benefit field-working linguists (section #).

**Extending Linguistic Experimentation to Under-Researched Languages**

Many (psycho)linguists have made strong claims about universals of language processing, acquisition, and representation (Altmann, 1997; Altmann & Gaskell, 2007; Ambridge & Lieven 2011; Eisenbeiss, 2009a, subm., Kess, 1992; Traxler, 2011; Traxler & Gernsbacher, 2011). Some of the claims about language acquisition have been evaluated in cross-linguistic corpus-based studies on language acquisition (see e.g. Bittner, Dressler, & Kilani-Schoch, 2003; Guo et al., 2009; Slobin 1985, 1986, 1992, 1996, 1997a, 1997b; Stephany & Voeikova, 2009). However, there have been far fewer cross-linguistic that used experimental methods and there is still comparatively little cross-linguistic experimental research. Initially, the focus of experimental (psycho)linguistics was on English and closely related Germanic languages like Dutch and German, but recently, more research on Romance, Semitic, and South East Asian languages has been carried out. However, there are still very few studies on African languages, native American, and Australian languages; and research on Indian languages has been quite limited and mostly restricted to Hindi and Tamil. This means that the sample of languages studied in (psycho)linguistic experiments so far has been rather small and typologically homogeneous. Thus, a much broader cross-linguistic empirical base is needed for the evaluation of claims about universals of language processing, acquisition, and representation. The advantages of a larger empirical base have become particularly clear in research on (i) universals of sentence processing, and (ii) universals of child-directed speech.
Universals of Sentence Production

Some claims of competing psycholinguistic models can only be tested on languages with particular linguistic properties or processes. This becomes particularly clear when one looks at models of sentence production (see e.g. Jaeger and Norcliffe, 2009 for discussion). For instance, research on languages like English has demonstrated that noun phrases whose referents are highly salient (animate and topical) tend to be realised in a high-ranking syntactic position and at the beginning of the sentence. For instance, an animate agent are typically realised as a sentence-initial subject of an English sentence (The man broke the table; see e.g. Bock & Warren, 1985). This observation has been given two competing explanations: The first explanation is based on the assumption that cognitive accessibility affects word order directly so that noun phrases with referents that are easily conceptually accessible are realised earlier in the speech production process than noun phrases with less accessible referents. The second explanation is based on the assumption that cognitive accessibility affects word order only indirectly, by affecting the mapping of noun phrases onto particular syntactic positions.

Distinguishing the two alternative explanations on the basis of English is difficult as English word order is rather fixed, which means that word order and syntactic position are highly interrelated, noun phrases in higher syntactic functions (e.g. subjects) appearing early in the sentence. Hence, if one wants to compare direct and indirect mapping accounts, one needs languages with more flexible word orders and a growing number of studies has done that; see e.g. Chang, Lieven, and Tomasello (2008), Jaeger and Norcliffe (2009), Köhne, Pickering, & Branigan (2013) for overviews of studies on Cantonese, Fijian, Finnish, German, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Odawa, Russian, Spanish, Slovenian, and Turkish. This cross-linguistic research has shown that referent accessibility can affect word order...
independently of grammatical function assignment and in typologically very different languages with different ways of realising noun phrases syntactically and morphologically. Such reliably observed direct effects of accessibility on word order challenge models of sentence production that postulate two strictly ordered stages – an earlier functional processing stage, in which referents are mapped onto syntactic functions, and a later positional processing stage, in which these functions are mapped onto linear positions (see e.g. Levelt, 1989).

While cross-linguistic studies with a wider empirical base have helped to evaluate competing models of sentence production, they have also challenged some assumptions about processing universals that had been uncontroversial for a long time. For instance, most models of sentence processing assume that speakers tend to order shorter and syntactically less complex constituents before longer and more complex constituents as short constituents can be formulated faster, and should thus be selected earlier for production (Ferreira & Dell, 2000; Hawkins, 1994); and studies on English and other Germanic language have provided ample evidence to support this assumption. For instance, it has been shown that English native speakers prefer the prepositional construction over the dative construction when the Goal argument is particularly long (I would only give my keys to neighbours who have never annoyed me in my life; Hawkins, 1994; Arnold et al., 2000; Bresnan et al., 2007; Wasow, 2002). Similarly, English speakers prefer the order Possessum < Possessor when the Possesor phrase is very long (e.g. the car of my neighbour who lives in the pink house at the end of the street; for discussion see e.g. Rosenbach, 2008). However, English is a head-initial language with verb-object order and prepositions; and studies on head-final languages with object-verb order and postpositions have not found the same short<long-preferences in speech production (see e.g. Dryer, 1980; Hawkins 1994; Yamashita and Chang 2001 for Japanese and Choi, 2007, for Korean). This has led some researchers to consider whether ordering preferences for short and long constituents are really universal or whether they can depend on the headedness of the respective language (Hawkins, 1994, 2004, 2007).
Universals of Child-Directed Speech

Studies on child-directed speech are another clear example of a domain where cross-linguistic and cross-cultural studies with a large and diverse sample of populations are needed to test claims about universality. While initial studies based on American mothers claimed that "motherese", the language used by these mothers to talk to their children, was syntactically simplified and contained many corrections, these claims had to be modified based on later studies. In particular, the studies reviewed by Gallaway & Richards (1994), Ingram (1989), Locke (1995), and O'Grady (1997) found that, the amount of direct interaction with children differed across cultures - and the ways in which parents interacted with their children could be shown to depend to which degree very young children already viewed as autonomous beings that were capable of interacting. Thus, widening the database seemed to challenge claims about universals. On the other hand, these studies provided evidence for at least some potentially universal properties of child-directed speech, in particular the preference for simple, but mostly correct sentences, slow speed with longer inter-utterance pauses, a high pitch and a high degree of variation in intonation. In addition, cross-linguistic and cross-cultural studies have demonstrated that one can capture the majority child-directed utterances using a very restricted set of frames like That’s a __ ; even in languages like Russian that have very flexible word order patterns (Stoll et al. 2009). Moreover, studies on English, German, Hindi, Russian, and Turkish have demonstrated that child-directed speech is characterized by a substantial proportion of "variation sets" or "variation sequences" like the one in (##), which is addressed to a two-year old English-speaking child (Eisenbeiss, 2003; Küntay & Slobin, 1996, Onnis et al., 2008, Slobin et al., 2011).

(##) a) let’s put J’s bottles in the refrigerator (Slobin et al., 2011)
b) want to put them in the refrigerator with me

c) let’s put J’s bottles in the refrigerator

d) we’ll put it in the refrigerator

e) let’s put it in the refrigerator

f) we’ll put it in the refrigerator

g) you can put it in

h) I’ll let you put it in yourself

i) you put it right in

j) you put it in there

k) put it right in the refrigerator

Such sequences of adult utterances show a constant communicative intention and different types of variation in form; e.g. lexical substitution and rephrasing, a shift from full noun phrases to pronouns, the addition, deletion or reordering of constituents; e.g. (Küntay & Slobin 1996, Eisenbeiss 2003, Slobin et al. 2011). These properties highlight recurring lexical elements and constituent boundaries, or the links between full noun phrases and pronouns. They can also offer evidence for word order flexibility and the optionality of particular constituent types. Moreover, they can highlight morphological contrasts as the sequence in ## demonstrates (Lieven, Salomo, & Tomasello, M. 2009; MacWhinney 2000).

(#)  *MOT: can you see now how the petals are closing up?  Thomas (2;01;22)

*MOT: can you see how these daisy petals?/.

*CHI: 0 [=l makes noises].

Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and(Resources_2013_10_06.doc 11
Other input properties that have been observed in studies on North American and European children could help children to recover from errors they might produce. For example, if a child produced a form like *singed* and the parent reformulated the child's utterance, using the target form *sang* that would show the contrast with the child’s non-target-like form. This type of reformulation could alert children to their deviation from the target and hence provide “implicit” negative evidence (Chouinard and Clark 2003, Farrar 1990, 1992). While experimental studies on English have suggested that learners might benefit from such input (Saxton 1997, Saxton et al. 1998, 2005, Valian and Casey 2003), more cross-linguistic and cross-cultural studies are needed to determine whether such reformulations are a universal feature of child-directed speech and whether they always have any positive effects on children's linguistic development.

**Extending Psycholinguistic Experimentation to Under-Researched Populations**

If one wants to evaluate claims about universals of language acquisition and processing, it is not sufficient to extend the respective studies to a broad range of typologically diverse and often under-researched languages. It is also important to investigate a broad range of different participant populations, with different educational and social backgrounds and different types of motivation to take part in experimental studies. So far, the majority of experimental studies in psycholinguistics and psychology been conducted with undergraduate students from comparatively rich Western countries, in particular America. Recently, this group of participants has been characterised as WEIRD - “western, educated, industrialized, rich, and democratic” (Henrich, Heine, & Norenzayan, 2010) and a growing number of researchers have questioned
whether results obtained from studies with this particular participant group can be generalised to the wider population in the West and beyond. On the one hand, this group is perceived as comparatively homogeneous, on the other hand some researchers argue that “American undergraduates are some of the most psychologically unusual people on Earth” (Henrich et al 2010, p. 29). Other researchers have pointed out that participants in experimental studies are typically signing up for experiments, either for course credit or because they are interested in the studies as they want to study or work in the same field themselves. Thus, this group of participants may not even be representative for educated Western people that all speak the same language and share the same cultural background; and several studies have tried to demonstrate this. For instance, Ahn, Ostrom, & Walker (2010) have shown that in economic studies, self-selecting undergraduates do not always behave in the same way as postgraduates and researchers who take part in the same experiment as part of a summer school.

Similar arguments have been made for language acquisition studies carried out within Western societies. Some studies explicitly focus on comparing children from different social and educational backgrounds. However, many studies that focus on potential universals of language acquisition do not involve such explicit comparisons. In principle, such studies should be carried out with a random sample of the population, but in practice and due to logistic reasons, many studies of this type are carried out in day care centres on a university campus or in other day care centres and schools that have high proportions of academics that can help researchers to obtain access for their research group. Other child acquisition studies are conducted in labs in universities or research institutions, with self-selecting parents who have responded to advertising on a university website or in day care centers – also not a randomly selected and mixed group.
Thus, even within rich Western societies and for seemingly well-studied languages, there is ample room for studies that compare the typical participant groups of psycholinguistic studies with different samples of the population. In response to this need, more and more researchers now investigate individual differences in learner's language attainment and native speakers' processing that could be related to gender, social class, input, and working memory (see e.g. Dabrowska 2012; Street & Dabrowska 2010, 2012; van den Brink et al., 2012 for overview). In order to carry out such studies, it is often necessary to extend psycholinguistic research to communities that are unlikely to sign up for experiments in a laboratory on a university campus and to use mobile labs to reach these communities, even for comparatively high-tech experiments such as eye-movement studies. When one attempts to extend psycholinguistic research to understudied population in this way, it is crucial to take into account those factors that have been shown to affect participants language learning and processing, in particular their gender, working memory, reading speed, language experience, familiarity with computers (and computer games). Thus, it is essential to collect relevant information about participants so that it can be taken into account for further analyses.

**Enriching Linguistic Fieldwork and Descriptive Linguistics with Experimental Methods**

Given the arguments discussed above, most psycholinguists can see the value of extending research beyond the small range of well-studied languages and population - although they might hesitate to explore this themselves due to a lack of fieldwork training or logistic and practical problems. Conversely, descriptive and theoretical linguists have recently become more open to experimental methods and the use of psycholinguistic data to support their arguments. This is in
part due to growing criticism of traditional linguistic research methods such as introspection and the elicitation of judgments from one or a few native speakers. Such methods involve comparatively subjective tasks - and they are also particularly vulnerable to researcher bias. This is obvious if linguists only rely on their own judgements about the acceptability, grammaticality or semantic interpretation of utterances and forms, but even working with one or a few other native speakers in a comparatively close social contact can lead to a situation where participants try to please researchers and provide the kinds of judgment that they would seem to prefer. Moreover, psychologists have demonstrated that acceptability or grammaticality judgments are affected by a range of factors that linguists have long not considered systematically enough – for instance length and working memory effects or the frequency of particular linguistic elements and their combinations in speakers' input (see Cowart 1998; Gibson & Fedorenko, 2010; Gilquin & Gries, 2009; Myers, 2009; Schutze 1996; Sprouse & Almeida forthcoming, and the special issue on experimental syntax, *Language and Cognitive Processes*, 2013). Such factors can be addressed using experimental methods when eliciting judgments. Similarly, more controlled elicited production experiments can often better identify factors that determine the choice of particular constructions, grammatical markers, or words than an elicitation session with a single participant alone could do (see e.g. Lüpke, 2010; Sakel & Everett, 2012).

Interestingly, the growing use of experimental methods in descriptive and theoretical linguistics has not simply increased the methodological repertoire of linguists. It has also led to more reflection about the effects of particular types of elicitation or even just the way an elicitation question is formulated or varied from participant to participant. Thus, adding experimental methods has led to improvements in the use of other research methods as well.

Despite such arguments in favour of experimental methods in descriptive linguistics, many
fieldworkers remain reluctant to include experimental tasks in their methodological repertoire. This is not simply due to old habits or a lack of training in such methods. After all, more and more linguistics departments around the world have made training in experimental methods obligatory. Some of this hesitation stems from legitimate concerns about the ecological validity of experiments. Most fieldworkers make deliberate efforts to keep the research situation as natural as possible to minimize the impact of the data collection situation on participants' responses. Hence, they are concerned about the use of strictly controlled experimental procedures and stimuli that are taken out of their discourse context and presented under time-pressure and often with technology that many participants are not completely familiar with.

In order to address such concerns, it is crucial to further develop and evaluate tasks that have already been shown to provide results that converge with those from more naturalistic studies (see the discussion below). Moreover, the use of experimental methods should be combined with naturalistic sampling and semi-structured tasks (potentially with stimulus materials) to provide converging evidence from a broad range of data sources. As will be discussed in section ## below, such more naturalistic methods are required in the stimulus preparation process anyway if one wants to conduct experiments for an under-researched language for which no lexical databases or other resources for stimulus-creation are available, yet.

Experimental methods cannot only be used to obtain more reliable and valid judgment data; they can also provide new types of evidence for fieldworking linguists and researchers who are interested in endangered languages, language documentation, or language maintenance. For instance, researchers who are trying to support language maintenance projects often find it difficult to compare the relative strength of the endangered language and other languages in the community and in individual speakers. One way to address this problem is a tool that draws on a
long tradition of psycholinguistic experimentation: the body-part naming task developed by the Hawai‘i Assessment of Language Access (HALA) project (O’Grady et al., 2009). This tool requires speakers to the same name body parts in each of their languages while response times are measured. The assumption is that shorter naming times show faster access of a lexical item and provide evidence for higher frequencies of use. In this way, comparing naming times for the two languages of a speaker should give us an indication of the relative usage frequencies and strength of the different languages that the respective speaker employs. This measure has been found to correlate highly with other measures of language strength and is easy to use even for illiterate speakers without a history of taking part in experiments. Note that psycholinguists can carry out further analyses of the resulting data by comparing naming times within the same language to obtain information about the relative frequency of words within the language. As will be discussed in section #, this type of information is crucial for the evaluation of claims about linguistic representation and for the development of stimulus materials for further experiments.

Thus, taken together, including experimental studies in linguistic fieldwork or documentation projects can provide new types of evidence for fieldworking linguists and researchers who are interested in language documentation or maintenance - while at the same time delivering insights into mechanisms of language development and language loss.

Making Use of Opportunities

As the discussion so far has demonstrated, taking experimental linguistics outside the traditional laboratory setting and applying it to under-researched languages and populations offers new insights and unique opportunities for psycholinguists. However, not all psycholinguists are themselves speakers of an under-researched language or associated with a particular fieldwork project.
site through established collaborations with fieldworkers. Thus, many psycholinguists who want to go beyond their laboratory will need support to find an appropriate language or field sites. Researchers who want to explore such options will find many resources that can help them to find appropriate languages, field-sites, collaborators and funding opportunities. A good starting point are introductory texts for fieldworkers that discuss crucial issues such as the availability of research on the language itself and related languages or the lingua franca spoken in the region (Abbi, 2001; Bowern, 2008; Crowley, 2007; Newman & Ratliff, 2001; Sakel & Everett, 2012; for updates, see http://experimentalfieldlinguistics.wordpress.com/readings/introductions-to-fieldwork/). Moreover, many web pages provide information about the specific linguistic properties of a broad range of typologically diverse languages around the world; e.g. the Ethnologue website (http://www.ethnologue.com/), the World Atlas of Language structures (http://wals.info/), and the typological databases provided by the Surrey Morphology Group (http://www.surrey.ac.uk/englishandlanguages/research/smg/webresources/).

While such resources might make it easier to make use of the opportunities that cross-linguistic research on under-researched languages has to offer, it is also obvious that the window of opportunity for such research is closing due to the rapid and continuing loss of languages and the fact that many languages are endangered to the point where they are only spoken by a very small group of people. While experimental studies are in principle possible for comparatively small populations, they are far more difficult to carry out with a limited pool of participants. At the same time the question arises whether the limited time and resources of fieldworking linguists would not be better spent on documenting an endangered language rather than on carrying out experiments with speakers of that language. Recall however that experimental approaches to linguistics can play a role even in a situation where language documentation and description and
not the evaluation of psycholinguistic models are the primary goal of linguistic research (see section #).

**Dealing with the Lack of a Fully Equipped Laboratory**

Experimental linguists typically work in a laboratory with specialist equipment and commercial, proprietary software and hardware for RT-measurements. Such laboratories offer a dedicated space with privacy and at least some sound insulation for other behavioural tasks, such as offline-experiments where participants have to describe or select pictures. Many laboratories also provide sound proof booths, multi-participant testing suites, and eye-tracking or brain-imaging equipment. For a long time, carrying out RT- and eye-tracking experiments required such an environment. However, this has changed due to advances in portable technology and initiatives for free and open-source software. These developments have made it possible to conduct RT-experiments and even some forms of eye-movement studies in other settings, using only free software and the types of equipment that linguists usually already have for their work. In section #, I will show how one can use "standard" computer and recording equipment for RT-measurements and some basic eye-movement studies. Section # will give an overview of free software packages and their uses in experimental linguistics.

**Equipment**

For RT-measurements, a standard PC or laptop is typically sufficient. For work with auditory stimuli or responses, one needs a good lightweight headset and a good built-in-microphone or – when phonetic analyses or voice-onset measurements are planned - a high-quality external microphone. All of these pieces are standard equipment for fieldworkers and many web pages provide up-to-date information about appropriate equipment (see e.g. Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc).
An additional piece of equipment that one might consider for RT-experiments is an input device that provides more accurate RT-measurements than a standard laptop keyboard. This could be an inexpensive and light-weight high-quality mouse or a USB-game pad. Alternatively, one can build such a device (http://reactiontimes.wordpress.com/arduinort/). The game pad has the advantage that it can be used with computer games that provide a source of entertainment for researchers and participants in a field site or in initial warm-up or reward-sessions with child participants.

Many fieldworkers employ a camcorder and some also use a tripod to record participants or to document the field site itself (see the links above for equipment advice). In some cases, researchers also carry a small projector with them as it can be useful in a field work setting, for instance for teaching or to show videos for elicitation or community entertainment. In combination with a laptop and a USB-game pad, the camcorder and the projector can be used to obtain basic low-resolution eye-movement measurements. Such measurements are sufficient in experimental paradigms where no reading is involved and it is only important to determine when a participant looks at a particular object or picture in a set of two to four objects or pictures. This is, for instance, the case in the visual world paradigm, which is commonly used to study the time-course of language comprehension processes in participant groups that do not read (see e.g. Clackson, Felser, & Clahsen, 2011. Huettig, Rommers, & Meyer, 2011; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Trueswell 2008). In such experiments, participants' eye-movements to individual objects or pictures in a display are measured while they listen to sentences or short narratives and look at the displays of objects or pictures, some of which were referred to in the
stimulus text. When participants are required to manipulate real objects, their actions can be analysed as well. For instance, several visual world studies have presented listeners with the sentence in (##) while they looked at a setup that involved (i) a frog sitting on a napkin, (ii) a second frog without a napkin, and (iii) an extra napkin plus another distractor object or animal. (see references above for discussion).

(##) Put the frog on the napkin into the box.

When listeners have reached the point when they hear the word napkin, they cannot definitely know, yet, whether the sentence ends at this point and they have to take the frog without the napkin and put it on the empty napkin. The sentence could also continue as in (##) , and they would have to take the frog that is sitting on the napkin and put it into the box. Monitoring participants' actions and/or their preparatory eye-movements allows us to see which frog they focus on when they have processed the sentence in (#) up to the word napkin. This gives us an indication of the interpretation they initially prefer; and many of the studies mentioned above have investigated which factors affect interpretation preferences and the course of processing (the use of different distractors, differences between child and adult participants, etc.).

For such experiments with simple displays, it is not necessary to measure eye-movements as finely as one would in an eye-tracking experiment on reading, where it is crucial to capture fixations on individual words or even letters. Hence, one can use "poor man's eye-trackers" (see e.g. Sekerina, Fernández, & Clahsen, 2008, in particular chapter 4). These can come in different versions, but all of them use a camera to record a video of participants' faces and eyes; and one can use free multi-media annotation software like ELAN (see section ## below) to annotate...
which object the participant focuses on during a given time-interval. If the experiment involves the display of pictures on a screen or as a projection on a wall or if participants have to press buttons to give answers to questions, free RT-measurement and presentation software can be used to control displays or record responses with appropriate timing accuracy (see section # below). For an overview of set-ups, see e.g. http://gameswithwords.fieldofscience.com/2010/02/lab-notebook-building-better-eyetracker.html: Three types are common and do not involve any helmets or chin rests, which makes them appropriate for children and for participants that are not used to laboratory equipment:

The first option is to use a laptop to display pictures, with a built-in camera to record eye-movements, a headset for the presentation of auditory stimuli, and a USB-game pad to record push-button responses (if required by the experimental design). The second option is to create a "box" with positions for pictures or objects in the four corners and a built-in-camera in the middle, which can be used for eye-movement recordings. This can be combined with a digital voice-recorder or a laptop with a headset to present audio-stimuli. Alternatively, audio stimuli can be spoken by the experimenter if a more naturalistic setting is desired for the task and a reduction of control over the sound quality and pronunciation of the stimulus can be accepted. Note that for work in a remote fieldwork setting, it might be advisable to have the box itself created in the fieldwork location to avoid logistic problems and to enhance community engagement with the project. The third option is to use a laptop, potentially with a USB-game pad to record push-button responses, and a data projector to project pictures on a screen. This can be combined with a camcorder on a tripod that is aimed the participant's face from below, see Figure 1 for picture of the participant, the camcorder and the screen presentation.
All of these setups are portable, though the first one is considerably more light-weight and less bulky than the other ones. The portable set-up in Figure 1 weighs around 20kg and is best transported in one large backpack or distributed between a trolley case, a bag for screen and tripod, and a backpack. Depending on the location, one can leave out the screen and the tripod and use a light-coloured wall for projections and a fixed box or another largish stable object to prop up the camcorder (similar to the propped-up projector in figure 1).

Figure 1: A Poor Man's Eye-tacker - Display, Camcorder and Participant with Headset

Table # provides a summary of essential and optional equipment for the portable psycholinguistics laboratory described above.
<table>
<thead>
<tr>
<th>Device</th>
<th>Uses</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Laptop                        | display of stimuli and RT-measurements, creation of stimuli, analysis of speech production | • operating-system dependent vs. platform-independent software  
• high-quality webcam, microphone, speakers and large RAM |
| game pad, mouse, push-buttons | measurement of button-press responses                                 | Reliable recording of button-responses and RTs                                                                                       |
| Camcorder                     | recording of speech, gesture, sign language, eye-tracking            | No hard drives or other avoidable moving parts to reduce sound emission                                                              |
| Tripod                        | stable positioning of camcorder and/or microphone (recording of stimuli/responses) | stability/weight trade-off, potentially requiring different tripods for different situations                                           |
| external microphone           | recording of speech                                                  | • only necessary if the sound quality of built-in microphones in laptops, camcorders or headsets is insufficient  
• choice of uni-directional or omni-directional microphone |
| Headset (with microphone)     | presentation of auditory stimuli (and recording of spoken responses) | noise-blocking systems recommended if weight is not a major issue                                                                   |
| rechargeable batteries and/or solar-panels | for camcorders, (some types of) microphones, laptops          | • for locations without mains electricity  
• reduction of humming risk for audio/video-recordings when they are not plugged into the mains |
| storage media                 | backup and data exchange                                             | external hard/flash-drives                                                                                                           |

Table #: A Portable Laboratory for RT- and Eye-Movement Studies

If more money is available and it is possible to transport a large trolley case in addition to any other required equipment, one can buy a portable eye-tracker that is precise enough to measure eye-movements during reading. Commonly used systems that provide more accurate measurements, but do not involve any helmets or chin rests are the Tobii systems (http://www.tobii.com/eye-tracking-research/global/research/linguistics/) and the SMI-systems Experimental_Linguistics_in_the_Field_OppORTunities_ChallEnGes_and_ResourCEs_2013_10_06.doc 24
As technology is changing rapidly, it is recommended to consult the equipment pages listed above. One can also make use of the question option of the LINGUIST List and its archives to obtain more information (http://linguistlist.org/).

**Software**

Most established psycholinguistic laboratories use commercial and proprietary software packages for the creation and presentation of stimuli (e.g. Presentation or E-Prime) and for the analysis of the resulting data (e.g. Excel, Matlab, SPSS, SAS). However, tighter budgets and recent initiatives to develop and promote the use of free and open-source software have made it possible to collect and analyse RT- and eye-movement measurements without expensive software, which is crucial for fieldworkers who want to minimise their budgets and to train others for future low-cost studies of their own. For RT-measurements, new software packages become available on a regular basis. Hence, before using them it is always worth to obtain reviews and suggestions from mailing lists like the Linguist List (http://linguistlist.org/), the mailing list associated with the conference for Architectures and Mechanisms for Language Processing (AMLAP; http://www.aamlap.org/aamlap-list.html, or the mailing list of the CUNY conference for Sentence Processing (http://lists.qc.cuny.edu/mailman/listinfo/sentproc ). Two free packages are commonly used and provide reliable measurements and community support when questions arise: Linger (http://tedlab.mit.edu/~dr/Linger/) has the advantage of being comparatively easy to use and platform-independent; i.e. it works on Mac, Windows and Unix systems. Linger was originally designed for performing reading, listening, and other sentence processing experiments, but offers some flexibility for the design of other experiments. However, if one wants more flexibility, especially for the use of auditory and picture/video
stimuli, the recording of spoken responses and remote testing, DMDX is currently a better choice ([http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm](http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm); [http://cnl.web.arizona.edu/dmdx.htm](http://cnl.web.arizona.edu/dmdx.htm)). However, this setup is limited to Windows machines.

Researchers who use "poor-man's eye-trackers" will need to annotate video recordings of participants' eye-movements to capture when they are looking at which corner or object of the display. This can be done with the free, XML-based multi-media annotation software ELAN, which can be used with Mac, Windows, and Unix systems and is hence popular with fieldworkers and child language for the transcription of spontaneous and elicited production data ([http://tla.mpi.nl/tools/tla-tools/elan/](http://tla.mpi.nl/tools/tla-tools/elan/)): This type of software allows researchers to time-link transcriptions or individual responses for language-production experiments to the respective parts of the corresponding audio/video-recording or to add annotations for eye-movements to a particular section of a video- or audio-file that was created using a "poor man's eye-tracker". Moreover, ELAN offers useful features such as sound wave display, slow play, loop mode, and a silence recogniser, which make transcriptions easier. ELAN can also import and export transcriptions and annotations to formats that psycholinguists and fieldworking linguists are using for further analyses of their data. In particular, ELAN interacts well with the free phonetic analysis package PRAAT ([http://www.fon.hum.uva.nl/praat/](http://www.fon.hum.uva.nl/praat/)) and the free tools shoebox ([http://www-01.sil.org/computing/shoebox/index.html](http://www-01.sil.org/computing/shoebox/index.html)) and toolbox ([http://www-01.sil.org/computing/toolbox/](http://www-01.sil.org/computing/toolbox/)) that many fieldworkers use to manage and analyse their data. ELAN can also export data in the CHAT-transcription format ([http://childes.psy.cmu.edu/manuals/chat.pdf](http://childes.psy.cmu.edu/manuals/chat.pdf); MacWhinney 2000) of the CHILDES database for child language data, a format that is also used by many psycholinguists who transcribe adult native speakers or language learners. CHAT transcripts can then be searched or semi-
automatically annotated using the CLAN-tools provided by CHILDES (http://childes.psy.cmu.edu/manuals/clan.pdf). Note, however, that ELAN exports will only ensure that header lines and names for transcription lines will have the appropriate format. If researchers want to be sure that repetitions, interrupted utterances, utterances with incomprehensible parts, etc. are considered or excluded from counts as they wish, they also have to use the CHAT transcription conventions for the marking of repetitions, interruptions, etc. Tutorials for DMDX, ELAN, and CHAT are provided on the download sites listed above, but there is also a short tutorial that focuses on the needs of researchers who want to conduct experiments in fieldwork situations: http://experimentalfieldlinguistics.wordpress.com/teaching/fieldexperiments_software/

Free software is also available for other parts of the research process: Open office spread-sheet and word processing software can be used for the creation of stimulus lists and scripts and for data entry (http://openoffice.download-insider.com/). For statistical analyses and for many types of corpus analyses, one can use R, which is free, very powerful and flexible and supported by a growing community of researchers, many of them from the psycholinguistics community (http://www.r-project.org/: Adler, 2009; Baayen, 2008; Crawley, 2012; Dalgaard, 2008; Everitt & Hothorn, 2011; Field, 2012; Gries, 2009a, b). Free software is also available for sound recordings, the manipulation of sound files for stimulus preparation, and for the analysis of speech recordings: for instance, software like VirtualDub can be used to lift audio-tracks of videos and store them as separate wav-files for further analysis or transcriptions in ELAN. Audacity (http://audacity.sourceforge.net/) and PRAAT (http://www.fon.hum.uva.nl/praat/) can be used to create auditory stimuli or analyse recordings of production experiments.
Many researchers who want to conduct linguistic experiments outside of a traditional laboratory go to a fieldwork site or to educational institutions or community centers to pilot their stimuli and experiments and to collect their data. However, others have already established close contacts to communities of speakers with sufficient internet access and (computer)literacy. In this case, it can be an option to use the internet to collect experimental data or to use online questionnaires to develop and pilot stimulus materials. This raises many methodological issues that go beyond this paper, but are intensively discussed in a growing number of publications (see e.g. Brand & Bradley, 2012; Joinson, 2007; Reips, 2002, 2013; Mason & Suri, 2012). Some experimental software packages can be used for remote testing. For DMDX, see: http://psv1.psych.arizona.edu/~jforster/dmdx/help/dmdxremotetestingoverview.htm, http://www.u.arizona.edu/~jforster/dmdx/help/dmdxremotemonitoring.htm; https://reactiontimes.wordpress.com/2011/05/30/how-to-setup-an-online-experiment-in-dmdx/).

For a flash-based alternative, see scriptingRT (http://reactiontimes.wordpress.com/scriptingrt/). There are also dedicated websites that allow users to conduct their own online experiments. For overview and links, see: http://psych.hanover.edu/research/exponnet.html, http://www.wexlist.net/, http://www.onlinepsychresearch.co.uk/, http://groups.inf.ed.ac.uk/webexp/). For online surveys many researchers employ Survey Monkey (https://www.surveymonkey.com/), which can be combined with scriptingRT. Other researchers use systems that allow researchers to make secure small payments for participants, like Amazon's Mechanical Turk, which is frequently used for questionnaire- or rating studies (https://www.mturk.com/mturk/welcome).

For more information and updates on free laboratory software and tools, see e.g. http://experimentalfieldlinguistics.wordpress.com/links/software/ and links posted there.
Practical Issues

While using free software and cheap and portable equipment can enable research outside the laboratory, researchers who want to conduct experiments in remote fieldwork settings often face additional challenges that may be more difficult to overcome. A "traditional" laboratory provides privacy and optimised light, sound, and temperature conditions for conducting experiments, something that is obviously difficult to obtain in a remote village in a very cold or hot and humid region. Some of these issues can be addressed by selecting a season where weather conditions are optimal and participants are not likely to be engaged in village activities that create distractions and noise. Moreover, the introductions to fieldwork and the websites supporting fieldworkers that were already mentioned in section # provide a lot of advice on protecting equipment and improving recording conditions. In particular, they recommend the use of waterproof equipment pouches with re-usable moisture-absorbing silicone packs, noise-blocking headphones, directional microphones that are less likely to pick up background noise, foam-shields or wind muffls ("dead cats") that protect microphones against wind, solar-panels that provide electricity and light for recordings at more quiet times in the evenings, etc. Some devices for recordings in less than optimal conditions can be easily and cheaply created by researchers, and it is worth checking on youtube for instructional videos (see e.g. these instructions for creating wind muffls or screens:  

http://www.youtube.com/watch?v=4HRBBBe9vp9U; http://www.youtube.com/watch?v=SudmkbL06zg )

In addition, it can be useful to keep tools ready in case of equipment problems or to attach equipment in unstable situations, for instance a basic multi-tool or Swiss army knife, duct tape, rubber bands, string, clamps, and super glue.
In order to provide privacy and protection from the weather and outside noise or glare, some researchers rent their own individual accommodation or have access to rooms in schools or churches where they can record with some privacy and in a reasonably quiet environment. However, where this is not possible, researchers might try to get additional shelter – either a simple tent-like construction that provides at least some protection from light, wind, and preying eyes, or a more stable hut. Note, however, that creating such private spaces is not without problems as Dan Everett explained in his introduction to fieldwork (Sakel & Everett, 2012: 88):

"Peter Ladefoged and I once commandeered a hut in a Pirahā village to carry out phonetic investigations and I simply asked people to keep the children and dogs away and to keep the general noise level as low as possible. They kindly accommodated us.

Like any field activity, selecting and developing a location for your work can have unintended consequences. At a later stage I built a small wooden structure about 100 metres outside the Pirahā village, raised above the ground, and screened in, with a lockable door. I discouraged people from looking in during sessions nad tried to allow in only the teachers working in a given session, rather than them and all their friends. The Pirahā do not mindn this. But a Brazilian government agencies investigating the activities of a 'gringo' in the area asked the Pirahās about this small structure when I was absent from the village. The Pirahā, in their nearly non-existent Portuguese, were able to communicate that I spent a lot of time in there and that few people were allowed in. The agency representatives wondered what sort of fiendish experiments I might be running out there. Finding out about this, I quickly visited the agency headquarters and gave them a full explanation."

Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc 30
Dealing with a Lack of Psycholinguistic Resources

Anybody working on an under-researched language will need to build up knowledge incrementally, first obtaining basic information about the linguistic properties of the respective language and then using this information to find out even more. Psycholinguists face additional problems as they have become used to working with resources that one cannot simply create in a few weeks of fieldwork before creating experimental materials, for instance databases with information about frequencies of words and their forms, (the strength and type of) their semantic relationships. However, the number of available resources is constantly growing and many websites or institutions offer experimental linguists some support for the creation of new resources. This will be discussed below for corpora, lexical and frequency information (section #) and (ii) relatedness measures for word pairs (section #). Practical issues involved in the creation of resources will be addressed in section #.

Corpora, Lexical Information and Frequency Measures

While researchers working on under-researched languages will not find the same wealth of corpus data as researchers studying English or other well-studied languages, the number of corpora is constantly growing and new corpus-based resources become available all the time, often free of access-charges. Many of these resources can be found via the following web pages: http://talkbank.org/, http://www.ldc.upenn.edu; http://www.sketchengine.co.uk; http://www.mpi.nl/resources/data; http://dobes.mpi.nl/; http://www.elar-archive.org/index.php, http://chilides.psy.cmu.edu; for updates, see http://experimentalfieldlinguistics.wordpress.com/experimental-materials/lexical_databases/.

Updates about corpora and other resources are also available to subscribers of mailing lists for Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc
(psycho)linguists, e.g. the Linguist List (http://linguistlist.org/), the IASCL Child Language Bulletin (http://www.iascl.org/clb), and the CHILDES mailing list (http://childes.psy.cmu.edu/tools/email).

Most of the corpus websites listed above also provide information and support for researchers who want to create their own corpora. Many of them also offer to host and disseminate corpora and related resources. Researchers who want to create their own corpus resources for future work on an under-resourced language can supplement traditional recordings and text collections by using internet chat rooms and other online texts in the respective language as the basis for corpus creation. Advice on methodological issues is readily available (see e.g. Uthus & Aha, 2013 and publications cited there).

Researchers working on well-researched languages do not only have access to corpora, they also have access to corpus-based resources, for instance lexical databases with combinations of orthographic, phonological, morphological, syntactic and frequency information (for links to lists of corpora and related lexical databases, see: http://experimentalfieldlinguistics.wordpress.com/experimental-materials/lexical_databases/).

Again, it is worth checking existing web pages with such resources for updates on recently added languages. Some databases provide RT-data for words that can be very useful for computer simulations and for the evaluation of claims about factors that determine RTs (e.g. the British, Dutch, and French Lexicon Projects (http://crr.ugent.be/programs-data/lexicon-projects; Keuleers, Lacey, Rastle, & Brysbaert, 2012; Keuleers, Diependaele, & Brysbaert, 2010; Ferrand, New, Brysbaert, Keuleers, Bonin, Méot, Augustinova, & Pallier, 2010). For child language research, many researchers are using word lists or parent report forms to assess language and communication skills in infants and young children, in particular the MacArthur-Bates Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc
Communicative Development Inventories (CDIs, [http://www.sci.sdsu.edu/cdi/index.htm](http://www.sci.sdsu.edu/cdi/index.htm)), which have adaptations for a growing number languages and guidelines for the creation of new CDIs ([http://www.sci.sdsu.edu/cdi/adaptations_ol.htm](http://www.sci.sdsu.edu/cdi/adaptations_ol.htm)).

For many psycholinguistic experiments, researchers need quantitative information about lexical elements, for instance about the frequency of particular word forms or about the degree of formal or semantic relatedness between them. If lexical resources are not available for the language under study, one needs to start out by compiling lists of potential stimuli. Existing word lists can be very helpful, even if they do not consist of words from the respective language. Word check lists for fieldwork elicitation are a good starting point for compiling a word list in the target language for the experiment (see e.g. Abbi, 2001; Comrie & Smith; Sutton & Walsh; Swadesh, 1952, 1955; [http://experimentalfieldlinguistics.wordpress.com/experimental-materials/wordlists/](http://experimentalfieldlinguistics.wordpress.com/experimental-materials/wordlists/)). For research on children, one can start with the MacArthur-Bates Communicative Development Inventories for related languages and similar cultures ([http://www.sci.sdsu.edu/cdi/index.htm](http://www.sci.sdsu.edu/cdi/index.htm)). Moreover, word lists from comparable experiments for other – ideally related – languages can provide a good starting point as well. Such lists are often provided in the appendix of the respective research article or directly from the authors of the publication.

Once initial word lists have been created on the basis of such resources, native speakers can be asked to check them for any properties that might affect responses in experiments, for instance lexical ambiguities (like *bank*: river bank vs. financial institution), grammatical ambiguities (e.g. *run*: noun vs. verb), unwanted connotations, or potential offensiveness. It is also crucial to ask native speakers of different age ranges and backgrounds about potential associations with events or entities of particular cultural significance that might distract participants. For instance, words
or sentences that are (part of) well-known songs, poems, or catch phrases in movies or TV-shows can amuse or otherwise distract participants and lead to longer RTs. Moreover, it is important to control for grammatical and morphological factors that can affect experimental results, such as idiosyncratic case-marking properties, irregular inflection, gender or noun class, etc. Thus, in a fieldwork context, it is crucial to plan data collection so that such information is collected first.

For studies on lexical processing, it is also important to check whether forms are pseudo-derived, i.e. look as if they contained an affix even though they are mono-morphemic, such as the word *corner*, which looks as if it were a combination of the stem *corn-* and the affix *–er*. In experiments on lexical processing and representations, such forms can behave differently from mono-morphemic words that do not contain strings that could be interpreted as affixes (see Kgold & Eisenbeiss, 2013, for discussion and references). Such "pseudo-derived" forms should only be included if pseudo-derivation is a topic of investigation for the respective study. Another factor is family-size, the number of different forms of the same word (De Jong, Schreuder, & Baayen, 2000). Finding out what the members of a word's lexical family are can be part of the elicitation process in a fieldwork context, but it is crucial that the elicitation is aimed at getting an exhaustive list of related forms and that it is scheduled before any experiments to ensure that family size is controlled for in the experiments.

Once a list of potential stimulus words is compiled and checked for the properties mentioned above, researchers can create questionnaires, which can be used (online) to obtain frequency information. As the frequency of words (and their combinations) leads to stronger memory traces, it is unsurprising that it affects language processing (for reviews of studies on a range of different populations, see e.g. Bybee, 2006; Ellis, 2002; Monsell, 1991). Hence, one needs to control for frequency. This poses a problem for studies on under-researched languages, where...
frequency databases or large corpora are not available. However, it has been demonstrated that subjective frequency ratings from a set of at least 20 native participants tend to correlate well with corpus-based frequencies and RTs for lexical decision experiments (Schreuder & Baayen, 1997, Kgolo & Eisenbeiss, 2012). For such ratings, participants are typically asked to rate the frequency of a given word form on a scale from 1 to 5 or 1 to 7. Note that many researchers prefer such scales to scales with a zero in the middle (e.g. a scale from -3 to +3) as participants often select the zero when they do not know what to answer and perceive zero as the default or "I don't know" answer. Recently, however, some researchers have argued that one can obtain better measurements by replacing scales with numbers by visual scales where participants have to mark points on a line, which can be done very easily in online questionnaires (see e.g. Funke & Reips, 2012 for discussion). However, so far, such scales have not been tested extensively for frequency ratings. Moreover, when ratings are obtained using online questionnaires (see section ## for tools), it is crucial that instructions are clear and also contain practice materials with examples. In particular, it is important to clarify whether participants are asked about the frequency of a particular form or about the frequency of a word in all of its forms. It is also important to look at standard deviations or other measures of variability for the responses to determine whether some words might be more familiar to some (groups of) speakers than others, maybe due to their professional or leisure activities.

Once frequency information is established, it can be used to create stimulus lists for experiments. Note that in many experiments, stimuli are only supposed to differ with respect to one or two variables (e.g. word form frequency), while being matched with respect to others (e.g. stem frequency or length). Matching large sets of stimuli with respect to several variables at the same time can be very time-consuming, but researchers can now use online tools like the Match tool to
speed up this process (http://www.mrc-cbu.cam.ac.uk/people/maarten.van-casteren/mixandmatch/).

Relatedness Measures for Word Pairs

When pairs of words are used as stimuli, researchers typically control or manipulate the degree and type of semantic relatedness between them. For instance, some experiments may require word pairs that can be considered synonyms; and researchers working on well-researched languages can rely on databases that help them to find such pairs, e.g. databases created by the global WordNet initiative (http://globalwordnet.org/). WordNet databases show semantic networks with different types of semantic relationships (e.g. synonymy, hyperonymy/hyponymy, meronymy, and antonymy). Such databases are available for a broad range of languages and the global WordNet initiative offers support for researchers wanting to create new WordNets.

Researchers working on word pairs typically also need more detailed quantitative information about the degree of semantic relatedness between two words in a pair, for instance to compare pairs with a closer semantic relationship to less semantically less closely related pairs of words. For languages like English, databases with relevant information are readily available (see e.g. Buchanan, Holmes, Teasley, & Hutchison, 2013 and references cited there). Moreover, researchers can often find pairs with particular types of semantic relations by using stimuli from appendices for earlier studies for which semantic relatedness ratings or other types of quantitative measurements for semantic relations have been conducted. For many other languages, there are often at least some printed or digital Thesaurus resources that allow researchers to find semantically related pairs of words for which semantic relatedness questionnaires can provide information about the degree of relatedness.

When such resources are not available for the language under study, researchers must obtain the Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc 36
relevant information using the same methods that were employed to created resources mentioned above – mostly semantic relatedness questionnaires and association studies. In practice, particularly in a fieldwork setting, researchers often have very little time for the preparation of experiments. Moreover, conducting a rating task is difficult when the researcher is not a native speaker and when no lexica are available to support the creation of lists for rating tasks.

In such situations, a step-wise process is required: First, an initial list of potential stimulus words has to be created using word checklists and further lexical elicitation tasks, as described above. Participants can then be asked for word associations. This process can be started informally part of the lexical elicitation process, but it should ideally be followed up with a combination of standardised association tasks: Discrete word association task, where participants produce a single response to a given word, tend to provide strongly associated words. In contrast, free association tasks, where participants are asked to give as many possible responses as possible to a given word, tend to provide information about less strongly related word pairs that are nevertheless semantically related. Both types of association tasks can either be done in situ or via an online questionnaire. In addition, researchers can elicit translations for lists of related and unrelated word pairs that are taken from studies or databases for other languages.

This initial process typically results in a very long list of potential stimulus word pairs that has to be split up into several lists for semantic relatedness ratings. Assignment to these lists should be random, but an effort should be made to control for those properties that are already known. For instance, one should not create one questionnaire that contains mostly noun-noun pairs and another questionnaire that contains an equal mix of noun-noun pairs and noun-verb pairs. Moreover, each word should only appear once in each questionnaire, though it is possible to obtain ratings for different pairings of the same word through a combination of several
questionnaires that are given to different groups of participants. Finally, different types of semantic relations should be included and distributed across lists (e.g. synonymy, hyperonymy/hyponymy, meronymy, antonymy, and context-based associations such as bread-butter).

Once semantic relatedness ratings or questionnaires about types of semantic relations have been obtained, they can in principle be used for the creation of experiments in which semantic relationships are either manipulated or controlled. One might, however, consider to first run a semantic priming study to further examine semantic relatedness effects: Such a study allows one to determine whether the naming or lexical recognition times for a word are reduced after the presentation of a word that is assumed to be closely semantically related (e.g. apple – pear), compared to target words with a prime that is assumed to be less closely semantically related (e.g. table – pear; for overviews see Neely 1991; Tulving & Schacter, 1990).

When one analyses the results of the association tests, relatedness rating questionnaires, and semantic priming experiments, it is also important to consider measures of variability for the responses to determine whether some word pairs might be more closely linked for some (groups of) speakers than for others, maybe due to different levels of exposure to related items over the years or to likelihoods of co-occurrence in the language they are exposed to in their professional or leisure activities. Thus, similar analyses should be applied as in the evaluation of frequency ratings, which was discussed above.

Note that when no previous semantic relatedness or priming studies are available, it might be better not to use a factorial design, where word pairs have to be assigned to categories like "strongly related" and "unrelated" – and potentially also "weakly related". Such a design requires very robust assignment criteria for the different categories; and the initial ratings might not be
optimal, especially if they were obtained from a small pool of participants who have not been involved in experiments before. If one uses a regression model, no categorical distinctions between levels of relatedness have to be made (see e.g. Baayen, 2008). Instead, reaction-times are predicted on the basis of quantitative information for individual word-pairs (e.g. about mean semantic ratings and/or the size of semantic priming effects from an earlier study). Moreover, if further quantitative semantic-relatedness measures become available at a later point – for instance through further rating studies – no-re-categorisation is necessary and one does not risk ending up with unequal numbers for the different categories. Rather, one can easily use the new information as a predictor in a revised regression model.

Note that the step-wise process for obtaining semantic relatedness information for experiments requires time and planning, the information obtained at each step can contribute considerably to the description of an under-researched language. Moreover, it can help to create resources that can be useful for the community, such as a Thesaurus.

When pairs of words are used as stimuli, researchers not only control or manipulate the degree and type of semantic relatedness between them; they also need to consider the degree and type of formal relatedness. For instance, some experiments on morphological processing explore whether priming effects that have been observed for pairs like walked-walk are simply due to the fact that the target letter string, the base-form of the verb walk, is fully contained in the letter string for the prime form, the past tense form walked (see # for discussion). Thus, researchers must consider the amount of formal overlap in their analysis, and they often add control conditions where word forms show formal overlap even though they are not morphologically or semantically related (e.g. car-card).
Thus, researchers who want to carry out priming studies must (i) have measures of formal overlap and (ii) construct pairs of words that are not morphologically related but show form overlap. While researchers working on well-studied languages can often already find stimulus lists or data bases with the relevant information, others can use free online tools such as the match calculator (http://www.pc.rhul.ac.uk/staff/c.davis/Utilities/MatchCalc/). This tool calculates a value between 0 and 1 that represents the orthographic overlap or "match" between two strings of letters, where 0 shows no match and 1 indicates a perfect match. For instance, the words *cap* and *cat* have an overlap value of 0.67 as two out of three letters match. This tool also offers the option to compute phonological matches, which is useful for work with auditory stimuli or illiterate participants.

When researchers want to construct pairs of words that are morphologically unrelated but overlapping in form and cannot rely on earlier lists of pairs like that, they can take their initial word lists and use scripts that take off final or initial letters - or both or scripts that change the order of letters. Such scripts can be written using free software like R.

**Practical Issues**

In order to save time and to ensure that experiments are well-designed and have appropriate stimuli, it is useful to plan data collection so that resources are created in an incremental way as part of the research. For instance, for production experiments, researchers need to find out more about the target structures, but they also need to establish good cues that will encourage speakers to produce these constructions. For this, they need to determine under which circumstances it is natural for speakers to produce the target structure, which aspects of the target structure can vary and which factors affect its use and specific realization. In order to achieve this without long waiting times before they design actual the experiment, it is often useful to first have
spontaneous interactions, then recorded conversations, followed by semi-structured elicitation tasks, while controlled experiments are left for the end of fieldwork. This will also help to obtain the necessary background information for the experimental study (e.g. frequency questionnaires, piloting of stimuli, etc.).

Note, however, that experiments do not necessarily have to be the final step in a study. Shat some experiments might require less background information and can be integrated into earlier stages of a study, where they can inform later work. For instance, picture naming tasks that can be employed to determine levels of language strength and language loss can be carried out as soon as good social connections with participants are established and it is possible to provide them with instructions they can understand easily. At the same time, the information that one can gain through this task is very useful as an early assessment task that allows researchers to group participants for later elicitation or experimental sessions.

Thus, there is no fixed and study-independent schedule for the integration of experiments into other components of a study. Researchers can get some guidance from the introductions to fieldwork that were mentioned above, but they should nevertheless create flow-charts and gant-charts to ensure that each piece of information that is required for further steps is available in time and that some time is planned in for potential delays or other problems.

**Working with Participants from Different Cultural Backgrounds**

In section #, I had discussed the opportunities that arise from working with under-researched languages and populations. However, this type of research also poses serious challenges. On the one hand, even standard stimulus materials that have been used in cross-linguistic studies on
well-researched languages in Western societies may not be appropriate for other participant
groups (section #). On the other hand, working on under-researched languages and populations
often requires researchers to adapt tasks to participants that are not literate or do not have any
experience with computers or experiments (section #).

**Stimulus Materials**

In cross-cultural and cross-population studies, there is a general tension between different goals:
One the one hand, researchers want to keep materials and tasks identical for all groups to ensure
optimal comparability of results. On the other hand, properties of the individual languages need
to be taken into account. For instance, in the large European COST action project on language
disorders in the multilingual world (#link bisli), materials for case marking elicitation had to take
into account that some languages only have default accusative and nominative case marking for
subjects and direct objects while other languages have some verbs with idiosyncratic dative
marking non the direct object (e.g. the German verb *helfen* 'to help'). Moreover, even amongst
languages with idiosyncratic dative marking, there is some overlap in verbs that show such
marking (mostly verbs of experience or social interaction), but it is impossible to find a
substantial set of verbs that have dative objects in all dative-marking languages. If one adds the
requirement that the verbs must be depictable and known by children, the set of verbs that can be
used for comparative work becomes quite small. This problem becomes even more pronounced
when one tries to adapt the materials to languages that are not as closely related as the European
languages in the original sample. Moreover, if the planned cross-linguistic study involves
participants from very different culture, it becomes more difficult to avoid offence or confusion
due to the use of culturally inappropriate materials. In order to ensure the appropriateness of
potential stimulus materials, any materials should always be checked by a native speaker and the
Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc
resulting set should then be piloted with some potential participants before the actual testing. The following aspects of materials need to be considered so that materials are as neutral, comprehensible, and inoffensive as possible:

Firstly, participants from non-Western countries may not be familiar with all of the Western conventions for picture, comics, and line drawings. Note also that the same is true for very young children who have not had much exposure to printed material, yet, and for Western participants who have not had much exposure to comics, e.g. participants from an older generation or some religious groups. For instance, many researchers have reported problems when they used drawings with central perspective or when they employed comic-style devices such as "thought bubbles" or wiggly lines that are supposed to indicate movement. Moreover, adult participants across cultures find the use of cartoons insulting as they consider them material for children. Hence, researchers have to determine whether the population in question is familiar and comfortable with comics and artistic devices and conventions. Even if this is the case for the current target population, such devices are best avoided in stimuli if one wants to have the option to use the stimuli for large-scale comparative studies later on. When it is possible to obtain them for the required types of objects, people, or actions, photographs are often a better choice of visual stimulus than drawings or cartoons.

Secondly, some animals, objects, professions, foods, etc. might be taboo, offensive, or provoke emotional responses that distract from the experiment. This is not just true for the pigs, magicians, sorcerers, witches, etc. that are commonly used in language acquisition studies in non-muslimic Northern and Western countries. It might also be a problem for people in more "everyday" professions. In particular, attitudes to people in uniforms vary considerably between cultures, which can affect responses. This creates a problem for cross-linguistic studies on
grammatical features like person, gender, case, or number. In such studies, researchers often want to depict easily recognisable animate event participants and use pictures of people in uniform instead of repeating general noun phrases like "the man", "the woman", "the boy", and "the girl". While such uses of uniforms can create strong emotional responses, it can also be confusing as uniforms for the same profession look different in different cultures.

Thirdly, it is important to consider norms and "in-group" rules for attire or the behaviour of males, females, and children. Note that this is important even within the same country, which is sometimes underestimated. For instance, some studies require people in stimulus materials to be clearly assigned to one gender; e.g. studies that investigate gender marking of pronouns or adjectives. American or European researchers then sometimes use drawings of women with heavy makeup, nail polish, skirts, figure-revealing clothing, and traditional female hairstyles; or they depict girls with highly gender-stereotyped frilly and pink clothing, nail polish, and jewellery. As I have experienced myself, this can lead to comments and distractions from the experiment when such pictures are used with participants that object to gender-stereotyping. Note that some of these problems can be avoided through the use of photographs instead of drawings as the sex of a person in a photograph is usually much easier to determine without special "markers" than the sex of a person in a drawing. Moreover, in cross-cultural studies, differences in attire for professions need to be considered as uniforms differ between countries and occasionally even between regions of the same country.

Participants can also be distracted from their task and focus on irrelevant aspects of the stimulus materials when the people that are depicted in stimulus photos, drawings or cartoons look different from the participants themselves with respect to skin and hair colour, hair style, height and size, dress, etc. While some differences cannot be avoided in cross-cultural work, they can
be minimized, e.g. by not asking the tallest and blondest student to pose for stimulus pictures and by avoiding the depiction of people with piercings, tattoos, and other specific or unusual properties.

Researchers who want to avoid offensive or culturally variable (depictions of) objects or professions, often use more abstract shapes. For instance, the caused-events video stimulus does not involve people or animals hitting or pushing each other, but circles and squares (reference and link). While this avoids some of the issues discussed above, it leads to rather unnatural comics, which some participants may find odd or childish. Moreover, care must be taken that participants can name the objects in the videos even if their language does not have names for abstract shapes like squares, triangles, or rectangles. With proper preparation, this can mostly be achieved by introducing objects and label them with available and appropriate nouns (e.g. "stick" for the rectangle).

Even when pictures involve familiar objects or animals and professions, participants may have experienced them in different combinations, depending on their culture. For instance, stimulus materials for acquisition studies in American or European labs often depict tigers, lions, and kangaroos together in picture series or even in the same picture, which matches the experience of American and European children who encounter these animals in zoos and often expect them all to live in the same "jungle". However, in the wild, these animals do not share habitats and their combination in picture materials can cause amusement or confusion when such pictures are shown to participants who have experience with these animals in their natural habitats.

Independently of cultural backgrounds, pictures should not involve distraction details even if that makes them more attractive. This is particularly important for eye-tracking studies, where patterns on a dress or eye-catching buttons or jewellery in a picture of a person might draw too
much attention. A simple “blink” test often already shows avoidable problems: If one closes one’s eyes, then briefly opens them for just one "blink”, any item or detail that immediately "sticks out”, may need to be removed or modified.

Researchers looking for pictures that have been used in different cultures, can find a growing number of resources. The best-known of them is probably the international picture database. This database is the result of a large international study to provide norms for timed-picture-naming in different languages and cultures (American English, German, Mexican Spanish, Italian, Bulgarian, Hungarian, and the variant of Mandarin Chinese spoken in Taiwan; http://crl.ucsd.edu/experiments/ipnp/). Thus, the pictures in this database have been tested for at least some different cultures. They can be searched by lexical parameters, percent name agreement, RT, and visual complexity. Another useful database is the International Affective Picture System, a database of photographs used in emotion research (http://csea.phhp.ufl.edu/media/iapsmessage.html).

Our website (http://experimentalfieldlinguistics.wordpress.com/) will provide further materials, for instance, a list of words that have already been used for work with children in a variety of cultures, with some basic information about gender or noun classes and transitivity of verbs. These words have also been translated and checked for appropriateness by adult native speakers of additional languages.

**Tasks, Experimental Designs, and Practical Issues**

When it comes to the selection of experimental tasks, illiteracy is not the only challenge that experimental linguists might have to face when they work on a broader range of languages, cultures, and populations. The greatest challenges is probably to find tasks that are not too alien and unnatural for participants who are not used to the strictly controlled nature of experiments or
to the use computers. While this challenge is expected by researchers who work in remote fieldwork settings where computers are not common, it also occurs when testing older and less educated participants in cultures where younger and more educated participants are familiar with computers and the rationale and nature of controlled experimentation. Moreover, difficulties in finding culturally appropriate tasks are not limited to experimental work. For instance, psycholinguists who want to collect spontaneous speech samples from different languages and cultures need to accept that it is not possible to compare mother-child play recordings from around the world: when mothers are engaged agricultural activities, outside paid work, or housework most of the day, they might not typically be the ones who play with their children (for discussion, see http://www.paradisec.org.au/blog/2012/10/researching-child-language-in-the-field-october-lip/). Rather, children might play with other – younger or older - children or grand parents might be their main adult interaction partners. Similarly, some of the games that language acquisition researchers use as semi-structured elicitation tasks, such as puzzles or picture-matching games, may be culturally alien activities for some of the populations they work with. This does not mean that children will not engage in them, but that playing the game will involve a certain novelty factor and the language they use might reflect that. Note that differences in common game types are not restricted to studies with languages that are spoken thousands of miles apart. For instance, Western societies differ with respect to the value they assign to imitation learning and games that involve children being asked to imitate what their parents or grandparents say. This may make elicited imitation tasks very easy and natural in some countries, while children need more practice with imitation tasks in other countries – even though they might in the end enjoy playing the parrot in a game. For instance, I have worked with some educated German who consciously avoid the types of imitation games that are
common in many Western societies as they are afraid that it might negatively affect children's own creativity and independence.

Psycholinguists constantly develop and evaluate new experimental tasks and paradigms (for overviews see Altmann & Gaskell, 2007; Blom & Unsworth, 2010, Huettig, Rommers, & Meyer 2011; MacDaniel, MacKee, & Cairns 1996; Sekerina, Fernández, & Clahsen 2008; Traxler & Gernsbacher 2011). However, researchers who want to extend current investigations to a new range of languages tend to prefer "classic" tasks that have been successfully used with a large number of languages and populations and can be carried out with minimal technical equipment. For illiterate participants, many researchers prefer to adapt established tasks from research fields where reading/writing is irrelevant, inappropriate, or a potential confound, e.g. research on sign language, child language acquisition, clinical populations, and phonology. Thus, textbooks and handbooks for research methods in these fields can provide a good starting point for the selection of methods (e.g. Ball, Perkins, Müller, & Howard, 2009; Blom & Unsworth, 2010; MacDaniel, MacKee, & Cairns, 1999; Menn & Ratner, 1999; Sekerina, Fernández, & Clahsen, 2008; Wei & Moyer, 2008). The overview in section ## below focuses on some of the tasks that have been successfully adapted to a broad range of languages, cultures, and populations.

Note that the extension of experimental research to understudied populations does not only require researchers to adapt stimuli and tasks. They also need to pay particular attention to the design of their studies, in particular if they want to carry out longitudinal studies or training studies that involve several sessions. Even in a more "traditional" laboratory setting, it can be difficult to find participants who are willing to take part in an experimental study, especially if there is no or little payment, the experiments are longer than half an hour, and the tasks are quite boring, such as reading and pressing buttons. Additional difficulties arise when the experimental
set-up requires participants to come back for an additional session. Most researchers deal with this issue by taking extra care with respect to advertising and participant selection. Many researchers also pay participants only (in full) after last session of the experimental series – or in spread-out instalments for longer studies. While these measures can also be applied when working with "non-traditional" participant groups, avoiding attrition in studies with such groups can be even more difficult. This is very obvious for nomadic cultures, where participants might not stay in the same place long enough to take part in a longitudinal study, but even in non-nomadic cultures many participants have work commitments that are incompatible with tightly scheduled and extended participation in a longitudinal study.

Additional issues arise in fieldwork situations where researchers interact with the population under study via "gatekeepers", who often have financial or political interest that must be taken into account (for a personal account, see e.g. http://moniyawlinguist.wordpress.com/2012/02/27/a-primer-on-what-linguistic-fieldwork-really-is-or-if-you-dont-think-about-culture-youre-screwed/ ). In such contexts, experimental linguists have to allocate additional time – and often additional financial resources – to include interactions with gatekeepers into the plan for their study. Note that interactions with gatekeepers can be even more difficult for experimental linguists than for "traditional" linguists as organizations or individuals who are already skeptical and suspicious about the motives and ultimate goals of outsiders when they themselves have little experience of experimental studies and are not familiar with their role and purpose. Thus, it is crucial to explain the value of experimental studies and the motivation of the researcher in sufficient detail before the study – and to de-brief not only the participants, but also the gatekeepers after the end of the study.
Adapting "Traditional" Experimental Tasks to "Non-Traditional Contexts"

The following section gives an overview of some established tasks that have been used with a comparatively wide range of populations already and are hence a good starting point for researchers who take experiments out of the "traditional" laboratory context.

Tasks for the Study of Lexical Representations and Processing

As mentioned above, of the most common task for the study of lexical representations and lexical access is the lexical decision task, where participants have to decide as fast as possible whether the stimulus is a word like *car* (YES) or a non-word/pseudo-word like *fonkel* (NO). In the visual lexical decision variant, the stimulus is a string of letters on a screen, while stimuli in the auditory lexical decision task are segments of spoken speech, which are presented via headphones or loudspeakers. The lexical decision task is highly sensitive to a variety of factors. In particular, the shorter and more frequent a word is, the faster it can be recognised as a word, which is attributed to strong memory traces (##references). This task has often been used to investigate how (morphologically complex) words are stored, with the following rationale: if a particular linguistic unit is stored, then the frequency of this unit should determine response times, in interaction with other factors such as length. Thus, if whole-word form storage played a role for complex forms like *walk-ed*, the frequency of the whole word form itself – and not the frequency of the stem and affix (*walk* and *ed*)– should affect RTs. In contrast, if a complex form is not represented as a unit, RTs should be determined by the frequency of its components and not by the frequency of the whole word form itself.
Auditory lexical decision tasks are an alternative to visual lexical decision tasks when participants cannot read, but this variant of the task poses additional challenges: The first challenge is obtaining good recordings of stimuli, avoiding potentially confounding factors such as background noise or inter-item differences in loudness or emphasis. This is time-consuming as one may need several trials per stimulus item to achieve this even in ideal laboratory contexts and with a trained and clearly articulating native speaker. In a fieldwork situation, this can be too difficult or time-consuming to achieve. Thus, the task may not be appropriate if the researcher does not have the time and language-specific knowledge to pre-select a broad range of potential stimuli and get a native speaker to-record them under laboratory condition before taking them to the field-site. The second challenge for auditory lexical decision tasks stems from the fact that an auditory stimulus is not presented "in one go", but as a sound stream that only becomes accessible over time. Thus, one has to be even more careful when it comes to potential competitors for the target that start with the same sounds (e.g. struggle and strung). The presence of such competitors with similar onsets could have a significant effect on response times. While such effects are well researched and can be factored in, this requires more attention to stimulus preparation. In particular, it is a very difficult and time-consuming to find potential lexical competitors when one does not already have lexical databases or corpora. In principle, though, eliciting potential competitors with similar onsets from native speakers can be part of the process of building up a lexicon for the new language as this process always involves looking for minimal pairs anyway.

An alternative task for studies on lexical representations and lexical is the naming task, where participants are asked to produce a word as fast as possible – either after seeing the word written on a screen or after seeing a picture that represents the word (## ref). Voice onset times are...
measured and the rationale for using these measurements them is similar to the rationale for using lexical decision times (but taking into account the different modalities involved). While the written naming task is obviously limited to literate participants, it has two advantages over the picture naming task: It can be used for target words that cannot be easily depicted (e.g. wisdom or cold), and both researcher and participant know what the desired output is. In contrast, with the picture naming task, different participants might use different words to describe the same picture. Note, however, that one can lower this risk by using pictures which have been extensively evaluated for different languages and cultures, for instance pictures from the international picture database, which was mentioned in section #(http://crl.ucsd.edu/experiments/ipnp/). For remote fieldwork settings, however, one might consider using photographs instead of drawings as they are more natural, as in the HALA task discussed in section #(#ref).

In order to investigate representations for linguistic units and their relationships, researchers have developed a range priming tasks. In all of these tasks, participants respond to a stimulus (=target) after the presentation of another stimulus (=prime); and it is assumed that the prior exposure to the prime can facilitate the response to the target by pre-activating the target – either directly or through an associative network. For instance, many studies have found that regularly inflected word forms like walked prime their respective stems (e.g. walk) just as effectively as the stem itself (see e.g. Pinker 1999). This has been taken as evidence for their morphological relationship and some researchers have argued that it suggests that regularly inflected word forms are decomposed into their stems and affixes, which could explain why these complex forms prime their stems just as strongly as the stems themselves. Priming experiments typically involve a lexical decision to a written target and a written prime or an auditory prime. However, one could
in principle also use an auditory prime and an ordinary target in a primed lexical decision task or one could use a picture naming task for the target and combine it with an auditory prime.

**Tasks for the Study of Grammatical Representations and Processing**

Researchers investigating syntactic processing typically use tasks where participants are exposed to sentences in an incremental fashion and their button-presses and/or eye-movements are measured to study (i) their interpretation of sentences, (ii) the time-course of processing or (iii) potential sources of processing difficulties (for overviews, see Traxler & Gernsbacher 2011).

One of the most commonly used tasks is the self-paced reading task. This task requires literate participants, but is very simple to implement as it does not require researchers to find appropriate pictures or record auditory stimuli. In this task, participants read through a text on the PC-screen in their own pace, by pressing buttons to consecutively display individual words or phrases on the screen. The display time for each word or phrase is measured to determine which parts of a sentence take longer to read, which indicates increased processing difficulty. The experiment may also contain some questions about the text. These questions give participants a reason to actually read the text – and they give you a chance to check whether they did this. The answers typically do not provide you with information about participants’ interpretation of the sentence. Reading experiments are obviously only appropriate for literate participants. Variants with audio-stimuli are possible in principle, but require even more control for the length of the individual segments and potential competitors for the words that are presented (similar to issues arising for auditory lexical decision experiments).

Researchers who work with illiterate participants or children often use the sentence-picture-matching task. In this task, participants see two pictures on computer screen, followed by a sound file with a sentence. They have to press a button to indicate which sentence matches the
picture. This task allows you to investigate both (i) the choice of picture, which provides information about participants' interpretation of the sentence, and (ii) the RT, which gives you information about processing difficulties. Thus, the sentence-picture-matching task gives the researcher more information about participants' interpretation of the sentence. However, as one only measures RTs for the entire sentence in a sentence-picture-matching task, one cannot tell at which point these processing difficulties occurred. Thus, in this way, this task provides less information than the self-paced reading task. However, if one compares response times for two sentences in the sentence-picture-matching task and finds significantly different response times even though the sentences only differ minimally, that is at only one point, one can easily determine the source of added difficulty that led to the longer response times that was observed for one of the sentences.

Researchers who are interested both in participants' interpretation of an utterance and in the time-course of sentence processing often use the visual world paradigm, where participants' eye-movements to objects or pictures are measured while they listen to sentences or short narratives and look at the objects or pictures (see section # above and Clackson, Felser, & Clahsen, 2011. Huettig, Rommers, & Meyer, 2011; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Trueswell 2008). When participants are asked to manipulate real objects, their actions can be analysed as well. For instance, when they have a frog sitting on a napkin and another frog and a napkin, one can see which of the two frogs they first pick up when they are asked to "put the frog on the napkin in the basket". When pictures are displayed on a screen, the display is controlled by RT-software like DMDX, and one can also measure RTs for button-press responses to questions about the stimuli. Both versions have been used in studies with adults and children, provide information about the time-course of processing, and are appropriate for illiterate participants.
Researchers who focus more on syntactic representations than on the time-course of processing, often use "structural" or "syntactic priming" tasks (Bock, 1986; Pickering, & Branigan 1998; Branigan, 2007). These tasks make use of the fact that participants are more likely to produce a given grammatical construction (e.g. a passive sentence like *The cat was chased by the dog*) after reading or hearing a different sentence with the same structure (e.g. the passive sentence *The window was cleaned by the man*) than after a sentence with a different structure (e.g. than after the active sentence *The man cleaned the window*). Such structural priming effects have been taken as evidence for abstract syntactic representations that can be pre-activated by priming.

Priming tasks can involve different types of primes (e.g. auditory vs. written) and target responses (e.g. lexical decision for a target word vs. production of a description for a picture or video in a structural-priming experiment), e.g.

While the discussion above has demonstrated how "classic" task can be carried out with minimal equipment and adapted to different types of populations, researchers should be prepared for more adaptations when they are working with participants who are unfamiliar with computers and recording equipment. In particular, for such participant groups, one might need additional training session before the actual experiment. These sessions can make use of tasks that are not directly related to the experiment, but involve the input device(s) for the experiment (e.g. the keyboard or the USB-game pad) and relevant types of displays (e.g. pictures, videos or sound, depending on the experiment). For instance, if one uses the projector setup in Figure 1 for eye-tracking experiments, it can be introduced as a projector for videos in a social setting first or one can first make video-recordings in naturalistic settings, without the projection of picture stimuli and without pointing the camcorder directly at the face of the participant. This can then be followed by semi-structured elicitation tasks that involve the presentation of stimuli on a...
computer, but in a more naturalistic, game-like setting instead of a strictly controlled experiment. Thus, the stepwise introduction of tasks that was recommended to build up language resources incrementally (see section #) can also be useful for training participants and familiarising them with equipment.

**Ethical Issues**

Any researcher carrying out experiments with human participants will have to meet national and international standards for such work. A link list for such guidelines and blogs associated with them can be found on our webpage for Experimental Linguistics in the Field ([http://experimentalfieldlinguistics.wordpress.com/links/linksethics/](http://experimentalfieldlinguistics.wordpress.com/links/linksethics/)). The key principles that are discussed in all of these guidelines are captured by the guidelines of the UK Economic and Social Research Council (ESRC; my highlights are mine):

(1#) **Key Principles of the ESRC Ethics Guidelines**

([http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf](http://www.esrc.ac.uk/_images/Framework-for-Research-Ethics_tcm8-4586.pdf), p.2f)

1. Research should be designed, reviewed and undertaken to ensure **integrity, quality and transparency**.

2. Research staff and participants must normally be **informed** fully about the purpose, methods and intended possible uses of the research, what their participation in the research entails and what risks, if any, are involved. […]

3. The **confidentiality** of information supplied by research participants and the **anonymity** of respondents must be respected.

4. Research participants must take part **voluntarily, free from any coercion**.
5. **Harm** to research participants and researchers must be avoided in all instances.

6. The **independence of research** must be clear, and any conflicts of interest or partiality must be explicit.

However, experimental linguistics outside a traditional laboratory context can raise specific issues, which will be discussed in more detail below. These issues arise from a unique combination of standards, ethical principles and legal requirements: Experimental linguistics must always meet the ethical standards for experimental linguistics, even if they take their experiments out of a traditional laboratory setting. However, if the experiments take place in a fieldwork context, they must also adhere to the constantly evolving standards for fieldwork and/or the standards for internet research. Moreover, psycholinguistic work with children, participants with cognitive impairments or other vulnerable populations requires additional considerations. Finally, experimental linguistics in a fieldwork situation is typically carried out in an international setting, which involves the exchange of data between different countries with different ethical guidelines and legal requirements. For instance, in the UK context, the Data Protection Act, the Freedom of Information Act, and the Human Rights Act need to be adhered to ([http://www.legislation.gov.uk.ukpga](http://www.legislation.gov.uk)). This complex interaction of requirements can lead to conflicts and tensions that need to be discussed in turn.

**Transparency, Quality, and Informed Consent**

Providing truly "informed" consent is difficult in a fieldwork setting, even before experimental methods come into play. This is particularly true for remote fieldwork sites where participants have little contact to the outside world and no internet access – and hence do not have the relevant background to completely understand the potential implications of making recordings.
available online. Hence, researchers working such situations must be careful not to exploit the
differences in background knowledge and persuade participants to share video footage online
when they have the feeling that their participants do not completely understand what that might
mean – and would object to the project if they did. For more discussion on this general problem,

While obtaining truly informed consent is in itself often difficult, the problems in obtaining
informed consent are compounded by tensions between the requirements for quality,
transparency, and informed consent. These tensions mainly result from the so-called "observer's
paradox" (Labov 1972:113), i.e. the fact that the objects of research themselves can be affected by
the observation that the research process requires. Thus, the research process itself can threaten the
very goals of the research project and adhering to ethical principles can enhance this danger: On the
one hand, researchers aim for ecological validity and want to keep the research situation as natural
as possible to reduce the impact of the data collection situation on participants' responses. On the
other hand, it is their duty to inform participants about the nature and aims of their research. It is
relatively unproblematic to provide details about those aspects of their experimental procedures
and materials that are likely to affect participants – e.g. which types of pictures or linguistic
materials they will see or hear, whether computer screens or auditory stimuli are involved,
whether there will be time-pressure as in reaction-time experiments, who will have access to the
data and any personal information collected as part of the study, etc. This information should not
only be provided for ethical reasons. Having it can also help participants to be prepared and
hence adapt better to the situation without any distractions. However, participants will also need
to be informed about the overall goals and aims of the research project, for instance, about the
researcher aiming to gain insights into the mechanisms of language processing or acquisition. Knowing these goals can enhance observer effects and lead to strategies, especially in an experimental context where it is more difficult to forget the artificial recording situation and the observing researcher than in the more natural conversational context of a spontaneous speech recording.

Providing information about research aims can lead to observer effects in any experimental research project. However, observer effects are an even bigger problem in a fieldwork setting where participants have not experienced experimental research before, but are more than familiar with outsiders coming into their community to exploit them or their natural resources for financial or political gain. In such contexts, participants are highly likely to have additional thoughts about the motivation of the researchers and spend at least some of their time in the experimental situation wondering about the researcher's ultimate goals. At the same time, it is more difficult to explain why one would like to conduct experiments than why one would like to learn the local language: "In my experience, questions about the researcher's motivations are among the most persistent and most difficult to deal with challenges to fairly negotiating the terms of collaboration with native speaker participants. This holds even for people you have known and worked with for decades. How much more, then, does it affect your work with participants you recruit for a single experiment and who you interact with for all of maybe 20 minutes!" (Juergen Bohnemeyer, p.c.).

At the same time, the fieldwork situation creates additional complication for obtaining informed consent and guaranteeing the transparency of the research process in experimental studies. For instance, on a traditional laboratory setting with (computer-)literate participants, researchers typically offer some debriefing after the experiment. In order to avoid participants sharing the
debriefing information before everyone has completed their tasks, this is typically done after all the data has been collected for this group of participants - either in a dedicated debriefing session for students in a class who all took part in an experiment, or via email, blogs, websites or posters in the respective department. In a fieldwork setting, this can be more difficult as data analysis might only be completed once the researchers have left the – potentially remote – field site and communication with all participants can be difficult. This is a particular concern when participants cannot read or lack computer access. However, even before data analysis is completed, researchers can provide at least some information about the more specific aims of the project and about predicted or initial results. Thus, a debriefing session before the end of the field trip is advisable.

Working with participants who are not literate poses another problem as information for participants and consent cannot be given in writing. In principle, information can be provided in spoken form and consent can be recorded using audio or video recordings. If participants have literate people in their community that they trust, a copy of the relevant documents can be left with them in addition to a copy of the recording. This will typically include a written consent form and a participant information form, potentially also a certificate showing ethical approval for the project. As in any project, these forms should be in plain language, without technical jargon or linguistic terminology. When consent is obtained orally, some researchers argue that "a spontaneous (if practiced) oral explanation of the project and consent issues is more understandable to people than a reading, even a reading of a very plain-language form, since it allows us to judge when and where people have questions or gaps in understanding and address them" (Peter Patrick, p.c.).
Note that recording oral consent raises additional issues as such recordings themselves can be seen as intrusive and participants who might agree to take part in an anonymous reaction-time experiment might not want to have their – clearly identifiable – voice or face recorded. Moreover, a protocol for obtaining consent has to be established and included in the forms submitted for ethical approval of the project. Hence, later ad-hoc modifications in the field are not an option and early planning and enquiries into potential problems with obtaining and recordings are necessary – something that experimental linguists working in laboratories have to adjust to. They also have to allow for situations where participants themselves can read and write in their own native language, but the researcher may not – yet – have the knowledge to provide information and consent forms in the participant's language. Then, the course of the study needs to be planned in a way that translators can be involved in the consent process.

Finally, obtaining consent in a fieldwork setting raises another issue that does not occur in a closed laboratory setting and requires decisions when consent procedures are developed: the comparable lack of privacy in many fieldwork settings. The following quote from an acquisition researcher working in a fieldwork setting illustrates the problem and describes a potential solution: "I have decided to keep access to my recordings of child language closed, until the children are 18. If they are happy for me to open access to their recordings after they are 18, I will do so. However since I am currently recording children in groups at least 3 people, it is likely that in many cases I will not be able to contact all participants so the recording will remain closed. One of the issues we returned to a number of times in the evenings is that our recordings are often made in open environments, which means that many people wander through the field of view. This is in contrast to mainstream child language data, which is usually made in a room through which only a limited number of people pass by. It was mentioned that the CHILDES
language database is a great example of an open access archive but it lacks much data from endangered languages. CHILDES contains data recorded from many different studies of child language acquisition. However to upload data to CHILDES you must have the consent of every person who appears, even if just walking past. This is not going to be possible for many recordings of endangered languages in remote areas. It is often difficult to find a room to record in and even if one is found, it is likely that many people will pass through it.” (http://www.paradisec.org.au/blog/2012/10/researching-child-language-in-the-field-october-lip/).

Another issue that arises when data are contributed to archives – but also in other cases - is the need to explicitly include a statement about future uses of data in consent, participant information, and ethical approval forms. Moreover, researchers should ideally determine beforehand whether some of the personal information that could be deleted after the current project could be valuable for future projects (e.g. on genealogy). In this case, this should be addressed in the ethical approval process and clear information about potential future uses of the data should be included in materials for participants – even if it is not clear whether such future projects will ever actually be funded or carried out. For some discussion of issues arising in the reanalysis of social science data, see Thompson (2000).

Confidentiality, Anonymity and the Data Protection Act

Both fieldworkers and experimental linguists have to reconcile increasingly strict rules for anonymity and confidentiality with emerging new standards in their field that require them to collect and record more and more personal information about participants. In particular, field workers are faced with the increasing meta-data requirements of archives and experimentalists have to consider a broad range of factors in the analysis of experimental data. This means that both types of researchers need to collect and store detailed information about participants’ age,
sex, educational, linguistic, and social background. At the same time, new issues arise for researchers who extend their work to different settings or use new methods.

Psycholinguists "in the field" might work in a "traditional" fieldwork setting or they might work in their own country, but with a small group of speakers of an under-researched language that they investigate outside the laboratory. In both contexts, they must address two issues: The first issue arises from the requirement that anonymization efforts should cover both names of people and names of places, institutions or landmarks that could lead to the identification of the population or the location of speakers and recordings. For well-known but small communities of speakers this may be very difficult; consider, for instance, e.g. the population of a well-known, but small island or specific small group of speakers that has a well-established and known expatriate community close to the university of the researcher(s). Note, however, that identification within a group is potentially more crucial for participants than the identification of the group itself, especially if recordings involve personal information that speakers are reluctant to share with other community members. In some cases, groups or institutions might even want to be publicly acknowledged as they are proud of their linguistic heritage or their contribution to research.

The second issue concerning anonymity is that the community of participants is smaller and less homogeneous than the usual groups of undergraduate from a large campus, who tend to be roughly the same age, with similar educational, linguistic, and social background. When you have a comparatively homogeneous sample from a very large population with similar characteristics, anonymity of results is comparatively easy to maintain, even if a lot of individual additional information about age, sex, linguistic, social, and educational is obtained from each participant. For instance, in a large group of linguistics undergraduates in a prestigious American
university there are bound to be several 19-year old females from a middle-class background that have a high-school degree and learned at least one foreign language as a second language. In contrast, for a small group in a fieldwork setting, the combination of age, sex, profession, social and educational background, etc. is often sufficient to uniquely identify a participant or at least narrow down the range of people to a very small number of participants, even if participants are given numbers or pseudonyms. This is particularly problematic for studies on children or participants with language impairments, where it is common practice to provide appendices that include information about participants' results in combination with relevant personal information, which should - in principle - be anonymised. Note that similar issues arise due to the current trend to provide more and more detailed meta-data for each fieldwork recording. Thus, these issues cannot be completely avoided in non-experimental work and researchers will need to find a balance between collecting and presenting relevant information about individuals and guaranteeing their anonymity.

Fieldworking linguists who want to add controlled experiments to their methodological repertoire are also faced with new challenges related to consent and anonymity. In particular, if they are working with participants who are unfamiliar with experimental methods, they will find it challenging to explain the implications of consenting to participate in an experiment. This is particularly true when experiments are combined with IQ-tests, working memory tests, and other standardized tests to control or study the role of factors like intelligence or working memory in language processing and acquisition. Participants who are unfamiliar with such tests have to be informed about the nature of these tests and their role in the research process, but they should also receive information about the role of such tests in other societies and the possibility to rank people on a scale based on their test results. This type of information is typically not included in
standard consent forms for laboratory-based studies in industrialized countries as researchers
who conduct studies in such a setting can usually assume that their participants are familiar with
the concept of controlled experiments and are aware of the implications of taking an IQ test or a
similar standardized test of cognitive or linguistics abilities. This is often different in a fieldwork
setting and simply describing the test procedures without mentioning the implications of test
results is ethically problematic. On the other hand, mentioning these implications might create
problems as participants might only start to attempt intra-group comparisons if they have been
made aware of the scalability of results. Thus, a careful balance has to be struck.

Even when participants are familiar with experiments and standardized tests and would in
principle consent to participate in a study involving them, some fieldwork settings that offer little
privacy may pose additional problems. Participants might very well be happy for their recordings
of a narrative being overheard by other people in the vicinity, but they might have different
views about tests that allow for a ranking within the community being overheard by others.
Moreover, even if participants can do the test in written format without the risk of being heard by
others, pressures to share information about test results in a small and close-knit community
must be considered before including such tests in the study. This is particularly important when
researchers have to rely on native speakers to help with data collection, transcription, and
analysis. In these cases, research assistants must not only be trained to use the relevant software
and data collection techniques, but also receive training in research ethics. While such
considerations are familiar to fieldworkers, they may be less familiar to laboratory-based
experimental linguists who usually work with post-graduate research assistants, who can
typically be relied on to have undergone both research methods and ethics training as part of
their own studies. Proposals for ethical approval should address such potential confidentiality issues and list strategies for managing them.

Ethical approval in the UK also involves a statement that ensures adherence to the Data Protection Act (http://www.legislation.gov.uk/ukpga). This act requires researchers (i) to only collect as much data as needed for their research, (ii) to only collect the type and amount of data that their participants have agreed to provide, and (iii) to not make the raw data or any personal information about individual participants available to other people unless this is explicitly stated in the signed informed consent form or recording provided by their participants. Other countries have similar types of legislation. This means that consent forms should be designed with potential future research in mind. At first glance, this might seem less relevant for experiments than for (semi) naturalistic speech recordings. The latter are typically analysed with respect to a broad range of research questions. Hence researchers are used to keeping the research goals stated in the consent forms comparatively broad and explicitly mention potential uses in further research projects and submissions to open-access archives. In contrast, experimental data are often only analysed for one publication and stored in closed archives. Hence, at first glance, this type of study seems less prone to data protection problems with further uses of data. However, this is slowly changing: On the one hand, new statistical techniques have led to a growing number of studies that involve meta-analyses of earlier data. Moreover, psycholinguists have started to submit reaction-time data to databases that allow researchers to run simulations and investigate effects of factors such as word form frequency, stem frequency, or length on word recognition times (see the discussion of the British, Dutch, and French Lexicon Projects above and http://crr.ugent.be/programs-data/lexicon-projects). As word recognition experiments are technically simple and do not require a lot of prior knowledge about the grammar of a given
language, obtaining such measurements might become an option for fieldworkers. It could in principle become part of their efforts to describe an under-researched language, in particular as part of the documentation of an endangered language. Thus, consent forms should allow for the later submission of anonymised data from experiments to be included in publicly available databases.

A final challenge for confidentiality, anonymity and future data use arises from the fact that many experimental studies in the field involve data sharing between different sites, often across country borders – something that is new to those experimental linguists who only carry out work in their own lab or with colleagues from the same country. While it is obvious that data sharing procedures and devices must conform to the regulations of all countries and institutions involved, it is often forgotten that most cloud-based services that researchers use for data sharing (e.g. Dropbox or google.docs) involve private companies and often servers in yet another country. Using university servers or international archives with virtual private networks (VPNs) may provide a solution, though upload/download times can be limiting for some data types (especially for audio/video-recordings). Traveling with data on laptops and hard-drives requires the use of password protection; and it may be advisable to separate personal information and data by storing them on different devices, ideally carried by different members of a research group or on a different trip. Moreover, the links between data files and files with personal information should not be immediately obvious from the filenames for someone unfamiliar with the study. Researchers should also consider which types of information actually need to be transferred. Often, the person analyzing the data in another country does not need names or dates of birth for participants, but only needs the results of the measurements together with more general information about the age, sex, educational, social and linguistic background.
Cooercion, Compensation, and Avoiding Harm

Both fieldworking linguists and laboratory-based experimental linguists have to consider the possibility that offering payments for participation in their studies could constitute coercion or unduly influence participants, distort their judgment, or compromise the voluntariness of their informed consent. On the other hand, participants should be fairly compensated for the time and effort they have to invest to take part in a study; and researchers want to express their respect and establish good working relations. In both fields, guidelines for compensation that take both considerations into account have been established and some general principles seem to emerge in the national and institutional guidelines mentioned above: Firstly, the amount of compensation should reflect the time and contribution and not constitute an undue inducement. Secondly, payment levels should be adjusted to payment levels in the respective community. They should be generous and fair without distorting local payment levels and discouraging participants from other work that is valuable for their community (e.g. teaching).

While laboratory-based experimental linguists tend to pay individual participants on an experiment-per-experiment basis, both fieldworkers and child language researchers have often found that it is better to contribute to the local community or host institution as a whole, for instance by buying equipment, teaching materials or educational toys. In smaller communities this may be a viable strategy for experimental linguists as well. When compensation is not made via payments, but in the form of goods, researchers are encouraged by many institutions and organisations to provide either locally resourced and ethically produced goods that support the local economy or to offer goods and services that cannot easily be accessed or paid for by the local community (e.g. technical equipment). Whether food or drinks can or should be offered to...
participants – especially for longer experiments that involve breaks – depends on the local situation and the guidelines of the institutions and organisations involved. In acquisition studies, there is a general trend away from sweets as treats and rewards towards educational toys for the community, school or day care centre or to simple ethically produced toys, stickers, or books for individual children. Individual recognition is in general more important if the time-investment for each child and their family has been considerable or if different members of the community have participated in many more experiments than others. However, in fieldwork situations that involve remote or relatively closed communities, such arrangements should be coordinated with any potential gatekeepers for the respective community. Moreover, experimental linguists who are starting to work in remote locations might need to consider monetary or other forms of compensation for local people (often expatriates) who are not participants in the study, but support their research and help them establish contact with potential participants.

Even if researchers ensure that their participants are not coerced into their experiments, they these to ensure that these participants do not come to any harm. Except for some neuro-imaging techniques that involve the injection of radio-active tracer materials, the types of experiments that linguists conduct are typically minimally invasive and carry out low risks. However, when participants are recorded while they reveal potentially sensitive personal information or political views, this could expose them to risks, in particular if data protection is not taken seriously enough. Such risks are typically lower in experiments than in naturalistic studies as experiments usually do not encourage participants to provide potentially dangerous information. However, as discussed above, taking part in standardized tests as part of an experimental study can potentially lead to problems in the community; and researchers should consider how they are going to avoid such problems and add relevant information to their ethical approval applications. Moreover,
some experimental methods have the potential to cause more stress to participants than others, such as experiments with negative feedback for slow reaction-times and incorrect responses, or paradigms where participants have to carry out more than one task at the same time so that task-interference effects can be measured, Hence, measures need to be taken to minimize stress, in particular when the study involves participants that have not been exposed to experiments before.

**Independence of Research**

The principle of independent research can be difficult to maintain in a "typical" fieldwork setting where linguists tend to have close personal relationships to the participants in their study. This makes it more likely that they do not just carry out research on these participants, but also for them ("advocacy research" in the terminology of Cameron et al., 1992). For instance, linguists often become involved in campaigns for educational or linguistic autonomy. Thus, researchers in a "typical" fieldwork setting, whether they carry out experiments or not, will have to find the right balance between maintaining the independence of their research and establishing supportive relationships with the participants in their studies. Many fieldworking linguists also aim to carry out research with their participants (carrying out "empowering research" in the terminology of Cameron et al., 1992). This approach requires the use of interactive and dialogic methodologies that minimise the objectification of participants in research studies. Note that this approach lends itself better to language documentation projects than to experimental studies, which have a built-in tendency to objectify participants. Thus, researchers who are aiming for an empowering research strategy will have to determine whether and how they can develop ways in which they can empower their participants in the research process when they carry out experimental studies. Researchers will also need to consider the potential impact of their findings on language policies. Being classified as an endangered language or as an independent language instead of a dialect...
can have serious consequences for funding and the provision of education in the respective language. Hence, linguists who might provide such classifications as part of their work can be under considerable political pressure from different interest groups and governments.

**Legal and Practical Issues**

In addition to the legal requirements resulting from data protection legislation (see discussion above), many institutions around the world require researchers and their assistants who want to conduct experiments to have police record checks, in particular for work with children or other vulnerable participants. Many research institutions or universities require such checks even if the families of participants, the local schools or other institutions do not consider them necessary – or might not have trust in police organisations due to their own experience. Obtaining police checks can cause problems in fieldwork contexts where a researcher may be working with a local research assistant. Many institutions have policies for situations where it is not possible to obtain police clearance for a local research assistant. In particular, researchers might have to obtain police checks for themselves and agree to supervise the local research assistants at all times during the experiment. Note that the situation in the UK has been changing continuously over the past decade and is likely to change in the near future, which makes it difficult to make any firm statements about this. This webpage provides the most current information (summer 2013) [https://www.gov.uk/government/publications/dbs-check-eligible-positions-guidance](https://www.gov.uk/government/publications/dbs-check-eligible-positions-guidance). One potential issue is the question of who will be entitled to request checks and who will be required to undergo such checks. While it is usually possible to find an appropriate solution for police check problems, this needs to be addressed and clarified before ethical approval is sought.

When several institutions are involved in the data collection process – for instance a university and a local school or university - ethical approval must sometimes be sought from both Experimental Linguistics in the Field Opportunities, Challenges and Resources 2013_10_06.doc 71
institutions. In other cases, the sending institution simply requires letters of approval and support for their own ethical approval procedure. As this depends on the regulations of individual countries and institutions and often changes from year to year, it is advisable to determine early on which procedures need to be followed. Moreover, time for potentially required multi-institutional reviews has to be planned in. Another factor that might lead to complications and delays is a lack of consistency with respect to the guidelines and forms that have to be filled in different countries. What is considered fair payment in one country might be thought of as undue inducement in another country. Moreover, in multi-institutional contexts, different labels (e.g. "payment" vs. "compensation") or box-headers in forms may seem a trivial matter, but in reality, they can make it difficult to match information or might lead to additional questions that require more time to answer and delay the approval process.

When experimental linguists go beyond their laboratories and use the internet to collect experimental data (see discussion above), additional ethical, legal and practical issues and dilemmas may emerge. These issues cannot be discussed in detail here as the focus is more on the combination of fieldwork and experimental linguistics. However, there is a growing number of publications and websites that researchers can consult, see e.g. Brand & Bradley, 2012; Joinson, 2007; Reips, 2002, 2013; Mason & Suri, 2012; the International Journal of Internet Research Ethics (http://ijire.net/) or the respective entry of the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/ethics-internet-research/).

Summary, further Readings, and Resources

Further readings and resources can be found on our resource web site (http://experimentalfieldlinguistics.wordpress.com) and on the resource and institutional web Experimental_Linguistics_in_the_Field_Opportunities_Challenges_and_Resources_2013_10_06.doc
sites listed there: http://experimentalfieldlinguistics.wordpress.com/links/institutes/. The Linguist (http://linguistlist.org/) provides access to further resources and other mailing lists. Experimental linguists will find list that are dedicated to conferences, new technical developments, summer schools, job and funding opportunities very helpful, for instance the mailing lists of the Cuny Conference on Human Sentence Processing (http://mailbox.qc.cuny.edu/mailman/listinfo/sentproc) and the mailing list of the AMLaP conference on Architectures and Mechanisms of Language Processing (http://amlap.coli.uni-saarland.de/amlap-list.html).
References

TBA

Appendix

TBA:

- Sample DMDX scripts
- Wordlists

Index

TBA
Acknowledgements

This text is linked to a resource website (http://experimentalfieldlinguistics.files.wordpress.com) and a handout that addresses some of the issues discussed here and provides software tutorials for DMDX, ELAN, and CLAN/CHAT (http://experimentalfieldlinguistics.files.wordpress.com/2013/09/eisenbeiss_et_alperimental_field_linguistics_dmdx_elan_chat_clan_2013.pdf). Naledi Kgolo, Elena Papadopoulou, and Kaili Clackson have contributed to the software parts of the tutorials and discussed relevant issues with me. Sarah Schmid and Janina Fickel have compiled references and links and other information for the website and the handout, which were the starting point for this text.

This article has benefited from collaborations and discussions with Anvita Abbi, Juergen Bohnemeyer, Birgit Hellwig, Ayesha Kidwai, Naledi Kgolo, Bettina Landgraf, Friedericke Luepke, Vaishna Narang, Peter Patrick, and Mandana Seyfeddinipur,